HUMAN REGULATORY PROTEINS

This application is a continuation application of U.S. application serial number 08/870,870, filed June 6, 1997, all of which applications and patents are hereby incorporated herein by reference.

FIELD OF THE INVENTION

This invention relates to nucleic acid and amino acid sequences of human regulatory proteins which are important in disease and to the use of these sequences in the diagnosis, prevention, and treatment of diseases associated with cell proliferation.

BACKGROUND OF THE INVENTION

Cells grow and differentiate, carry out their structural or metabolic roles, participate in organismal development, and respond to their environment by altering their gene expression. These functions are controlled by the timing and amount of expression attributable to thousands of individual genes. The regulation of expression is metabolically vital in that it conserves energy and prevents the synthesis and accumulation of intermediates such as RNA and incomplete or inactive proteins when the gene product is not needed.

Regulatory proteins are an absolutely essential component in the control of gene expression, turning individual or groups of genes on and off in response to various inductive mechanisms of the cell or organism; acting as transcription factors determining whether or not transcription is initiated, enhanced, or repressed; and splicing transcripts as dictated in a particular cell or tissue. Although these proteins interact with short stretches of DNA scattered throughout the entire genome, most gene expression is regulated near the site at which transcription starts or within the open reading frame of the gene being expressed. The regulated stretches of the DNA can be simple and interact with only a single protein, or they may require several proteins acting as part of a complex to order to regulate gene expression.

The double helix structure and repeated sequences of DNA create external features which can be recognized by regulatory molecules. These features are hydrogen bond donor and acceptor groups, hydrophobic patches, major and minor grooves, and regular, repeated stretches of sequence which cause distinct bends in the helix. These features provide recognition sites for the binding of regulatory proteins. Typically, these recognition sites are

less than 20 nucleotides in length although multiple sites may be adjacent to each other and each may exert control over a single gene. Hundreds of these DNA sequences have been identified, and each is recognized by a different protein or complex of proteins which carry out gene regulation.

The regulatory proteins or complexes recognize and bind to the following specific DNA sequences: 1) to the upstream (5') nontranslated region preceding the first translated exon of the open reading frame (ORF), 2) to introns which occur between the many exons of the ORF, and 3) to the downstream (3') untranslated region following the ORF. The regulatory protein surface features are extensively complementary to the surface features of the double helix. Even though each individual contact between the protein(s) and helix may be relatively weak (hydrogen bonds, ionic bonds, and/or hydrophobic interactions); the 20 or more contacts occurring between the protein and DNA result in a highly specific and very strong interaction.

Families of regulatory proteins

Many of the regulatory proteins incorporate one of a set of DNA-binding structural motifs, each of which contains either α helices or ß sheets and binds to the major groove of DNA. Seven of these structural motifs are helix-turn-helix, homeodomains, zinc finger, steroid receptor motif, ß sheets, leucine zipper, and helix-loop-helix.

The helix-turn-helix motif is constructed from two α helices connected by a short chain of amino acids, which constitutes the "turn". The two helices interact with each other to form a fixed angle. The more carboxy-terminal helix is called the recognition helix because it fits into the major groove of the DNA. The amino acid side chains recognize the specific DNA sequence to which the protein binds. The remaining structure varies a great deal among the regulatory proteins with this motif. The helix-turn-helix configuration is not stable without the rest of the protein, and will not bind to DNA without other peptide regions providing stability. Other peptide regions also interact with the DNA, increasing the number of unique sequences that this motif can be used to recognize.

Many of these and other sequence-specific DNA binding proteins actually bind as symmetric dimers to DNA sequences that are composed of two very similar half-sites, which are also arranged symmetrically. This configuration allows each protein monomer to interact

in the same way with the DNA recognition site, and to double the number of contacts with the DNA. This doubling of contacts greatly increases the binding affinity while only doubling the free energy of the interaction. This motif always binds to DNA that is in the B-DNA form.

The homeodomain motif is found in a special group of helix-turn-helix proteins that are encoded by homeotic selector genes, so called because the proteins encoded by these genes control developmental switches. For example, mutations in these genes cause one body part to be converted into another in the fruit fly, Drosophila. These genes have been found in every eukaryotic organism studied, including humans. The helix-turn-helix region of different homeodomains is always surrounded by the same structure (but not necessarily the same sequence), such that the motif is always presented to DNA the same way. This helix-turn-helix configuration is stable by itself and when isolated, can still bind to DNA. It may be significant that the helices in homeodomains are generally longer than the helices in most HLH regulatory proteins. Portions of the motif which interact most directly with DNA differ among these two families. Examples of DNA-protein binding are described in more detail in Pabo, C.O. and R.T. Sauer (1992; Ann. Rev. Biochem. 61:1053-95).

A third motif incorporates zinc molecules into the crucial portion of the protein. Although their structure can be one of several types, these proteins are most often referred to as having zinc fingers, based on early discoveries. Proteins in this family often contain tandem repeats of the 30-residue zinc finger motif, including the sequence pattern Cys-X2 or 4-Cys-X12-His-X3-5-His. These regulatory proteins have an α helix and an antiparallel ß sheet. Two histidines in the α helix and 2 cysteines near the turn in the ß sheet interact with the zinc ion, which then holds together the α helix and the ß sheet. Contact with the DNA is made by the arginine preceding the α helix, as well as the second, third, and sixth residues of the α helix. This arrangement is often found to be repeated as a cluster of several fingers, such that the α helix of each can contact the major groove of the DNA. This results in a very strong and specific protein-DNA interaction. By changing the number of zinc fingers, the strength of the binding can be altered.

The steroid receptors are a family of intracellular proteins that include receptors for steroids, retinoids, vitamin D, and thyroid hormones, as well as other important compounds.

The DNA binding domain contains about 70 residues, eight of which are conserved cysteines. This motif forms a structure in which two α helices are packed perpendicularly together, forming more of a globular shape than a finger. Each helix has a zinc ion near the start that holds a peptide loop against the N-terminal end of the helix. The first helix fits into the major groove of DNA, and side chains make contacts with edges of the base pairs in the DNA. The second helix of each monomer contacts the phosphate groups of the DNA backbone and also provides the dimerization interface. As do the helix-turn-helix proteins, these proteins form dimers that bind the DNA. In some cases, multiple choices exist for heterodimerization which produces another mechanism for fine-tuning the regulation of numerous genes.

Another family of regulatory proteins uses a motif consisting of a two-stranded antiparallel ß sheet rather than an α helix to recognize the major groove of DNA. The exact DNA sequence that is recognized depends on the amino acid sequence in the ß sheet from which the amino acid side chains extend and contact the DNA. In two prokaryotic examples of this family, the regulatory proteins form tetramers when binding DNA.

The leucine zipper family also forms dimers and has a 30-40 residue motif in which two α helices (one from each monomer) are joined together to form a short coiled-coil. The helices are held together by interactions among hydrophobic amino acid side chains (often on heptad-repeated leucines) that extend from one side of each helix. Beyond this, the helices separate and each basic region contacts the major groove of DNA. The overall configuration resembles a clothespin (the dimer) pinching a clothesline (DNA helix). Proteins with the leucine zipper motif can also form either homodimers or heterodimers extending the specific combinations available to activate or repress expression. This strategy is sometimes discussed as combinatorial control.

Yet another important motif is the helix-loop-helix, which consists of a short α helix connected by a loop to a longer α helix. The loop is flexible and allows the two helices to fold back against each other. The α helices bind both to DNA and to the HLH structure of another protein. The second protein can be the same (producing homodimers) or different (producing heterodimers). Some HLH monomers lack sufficient α helix to bind DNA, yet they can still form heterodimers. When a heterodimer is bound to an untruncated HLH, it is unable to bind DNA and serves as a mechanism to inactivate specific regulatory proteins.

Hundreds of regulatory proteins have been identified to date and more are being characterized in a wide variety of organisms. Most of them have one of the common structural motifs for making contact with DNA, but there are several important regulatory proteins, such as the p53 tumor suppressor gene, that do not share their structure with other known regulatory proteins. Novel variations on the known motifs and new motifs have been and are currently being characterized (Faisst, S. and S. Meyer (1992) Nucl. Acids Res. 20: 3-26).

Although binding of DNA to a regulatory protein is a very specific fit, there is not a universal code which matches a certain nucleotide pair to a certain amino acid. Hence, there is no way to predict the exact DNA sequence to which a particular regulatory protein will bind or the primary structure of a regulatory protein for a specific DNA sequence. Thus, the interaction of DNA and regulatory proteins is not limited to the area of the motifs described above. Other domains of the proteins often form crucial contacts with the DNA, and accessory proteins can provide important interactions which may convert a particular protein complex to an activator or a repressor, or completely prevent binding (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing Co, New York, NY pp.401-74).

Diseases and disorders related to gene regulation

Many neoplastic growths in humans can be traced to problems of gene regulation. Malignant growth of cells may be the result of excess transcriptional activator or loss of a suppressor (Cleary ML (1992) Cancer Surv. 15:89-104). Alternatively, gene fusion may produce chimeric loci with switched domains, such that the level of activation is no longer correct for the gene specificity of that factor.

The cellular response to infection or trauma is beneficial when gene expression is appropriate. However, when hyper-responsivity or another imbalance occurs for any reason, improper or insufficient regulation of gene expression may cause considerable tissue or organ damage. This damage is well documented in immunological responses to allergens, heart attack, stroke, and infections (Harrison's Principles of Internal Medicine, 13/e©, (1994) McGraw Hill, Inc. and Teton Data Systems Software©). In addition, the accumulation of somatic mutations and increasing inability to regulate cellular responses is seen in the prevalence of osteoarthritis and onset of other disorders associated with aging.

The discovery of new human regulatory proteins which are important in disease development and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention and treatment of diseases associated with cell proliferation, particularly immune responses and cancers.

SUMMARY OF THE INVENTION

The invention features substantially purified polypeptides, human regulatory proteins, collectively referred to as NHRP and individually referred to as NHRP-1, NHRP-2, NHRP-3, NHRP-4, NHRP-5, NHRP-6, NHRP-7, NHRP-8, NHRP-9, NHRP-10, NHRP-11, NHRP-12, NHRP-13, NHRP-14, NHRP-15, NHRP-16, NHRP-17, NHRP-18, NHRP-19, NHRP-20, NHRP-21, NHRP-22, NHRP-23, NHRP-24, NHRP-25, NHRP-26, NHRP-27, NHRP-28, NHRP-29, NHRP-30, NHRP-31, NHRP-32, NHRP-33, NHRP-34, NHRP-35, NHRP-36, and NHRP-37, having at least one of the amino acid sequences selected from the group consisting of SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19, SEQ ID NO:20, SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29, SEQ ID NO:30, SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, and SEQ ID NO:37.

The invention further provides isolated and substantially purified polynucleotide sequences encoding NHRP. In a particular aspect, the polynucleotide is at least one the nucleotide sequences selected from the group consisting of SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61,

In addition, the invention provides polynucleotide sequences, or fragments thereof, which hybridize under stringent conditions to the polynucleotide sequences of SEQ ID NOs:38-74. In another aspect the invention provides compositions comprising isolated and purified polynucleotide sequences of SEQ ID NOs:38-74 or fragments thereof.

The invention further provides a polynucleotide sequence comprising the complement or fragments thereof of any one of the polynucleotide sequences encoding NHRP. In another aspect the invention provides compositions comprising isolated and purified polynucleotide sequences comprising the complements of SEQ ID NOs:38-74, or fragments thereof.

The present invention further provides an expression vector containing at least a fragment of any one of the polynucleotide sequences of SEQ ID NOs:38-74. In yet another aspect, the expression vector containing the polynucleotide sequence is contained within a host cell.

The invention also provides a method for producing a polypeptide or a fragment thereof, the method comprising the steps of: a) culturing the host cell containing an expression vector containing at least a fragment of the polynucleotide sequence encoding an NHRP under conditions suitable for the expression of the polypeptide; and b) recovering the polypeptide from the host cell culture.

The invention also provides a pharmaceutical composition comprising a substantially purified NHRP in conjunction with a suitable pharmaceutical carrier.

The invention also provides a purified antagonist which decreases the activity of an NHRP. In one aspect the invention provides a purified antibody which binds to an NHRP.

Still further, the invention provides a purified agonist which modulates the activity of an NHRP.

The invention also provides a method for treating or preventing a cancer associated with the decreased expression or activity of NHRP comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition containing NHRP.

The invention also provides a method for treating or preventing a cancer associated with the increased expression or activity of NHRP comprising administering to a subject in need of such treatment an effective amount of a pharmaceutical composition containing NHRP.

The invention also provides a method for treating or preventing an immune response associated with the increased expression or activity of NHRP comprising administering to a subject in need of such treatment an effective amount of an antagonist of NHRP.

The invention also provides a method for stimulating cell proliferation comprising administering to a cell an effective amount of NHRP.

The invention also provides a method for detecting a polynucleotide which encodes a human regulatory protein in a biological sample comprising the steps of: a) hybridizing a polynucleotide sequence complementary to the human regulatory protein to nucleic acid material of a biological sample, thereby forming a hybridization complex; and b) detecting the hybridization complex, wherein the presence of the complex correlates with the presence of a polynucleotide encoding the human regulatory protein in the biological sample.

The invention also provides a microarray comprising one or more oligonucleotides derived from at least one of the polynucleotide sequences, SEQ ID NOs:38-74, encoding NHRP. In a particular aspect, the polynucleotides of the microarray are the polynucleotide sequences of SEQ ID NO:38, SEQ ID NO:39, SEQ ID NO:40, SEQ ID NO:41, SEQ ID NO:42, SEQ ID NO:43, SEQ ID NO:44, SEQ ID NO:45, SEQ ID NO:46, SEQ ID NO:47, SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, SEQ ID NO:51, SEQ ID NO:52, SEQ ID NO:53, SEQ ID NO:54, SEQ ID NO:55, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:58, SEQ ID NO:59, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:62, SEQ ID NO:63, SEQ ID NO:64, SEQ ID NO:65, SEQ ID NO:66, SEQ ID NO:67, SEQ ID NO:68, SEQ ID NO:69, SEQ ID NO:70, SEQ ID NO:71, SEQ ID NO:72, SEQ ID NO:73, and SEQ ID NO:74

The invention also provides a method for the simultaneous detection of the levels of expression of polynucleotides which encode human regulatory proteins in a biological sample comprising the steps of: a) hybridizing the microarray of oligonucleotides to labeled complementary nucleotides of a biological sample, thereby forming hybridization complexes; and b) quantifying expression, wherein the signal produced by the hybridization complexes correlates with expression of particular polynucleotides encoding human regulatory proteins in the biological sample. In a preferred embodiment, prior to hybridization, the nucleic acid material of the biological sample is amplified and labeled by the polymerase chain reaction.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, arrays and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

NHRP, as used herein, refers to the amino acid sequences of substantially purified NHRP obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist", as used herein, refers to a molecule which, when bound to NHRP, increases or prolongs the duration of the effect of NHRP. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to and modulate the effect of NHRP.

An "allele" or "allelic sequence", as used herein, is an alternative form of the gene encoding NHRP. Alleles may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding NHRP as used herein include those with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent NHRP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding NHRP, and improper or unexpected hybridization to alleles, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding NHRP. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent NHRP. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of NHRP is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

"Amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules. Fragments of NHRP are preferably about 5 to about 15 amino acids in length and retain the biological activity or the immunological activity of NHRP. Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, amino acid sequence, and like terms, are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "antagonist" as used herein, refers to a molecule which, when bound to NHRP, decreases the amount or the duration of the effect of the biological or immunological activity of NHRP. Antagonists may include proteins, nucleic acids, carbohydrates, or any other molecules which decrease the effect of NHRP.

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, $F(ab')_2$, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind NHRP polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal can be derived from the translation of RNA or synthesized chemically and can be conjugated to a carrier protein, if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin and thyroglobulin, keyhole limpet hemocyanin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "antigenic determinant", as used herein, refers to that fragment of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense", as used herein, refers to any composition containing nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules include peptide nucleic acids and may be produced by any method including synthesis or transcription. Once introduced into a cell, the complementary nucleotides combine with natural sequences produced by the cell to form

duplexes and block either transcription or translation. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic NHRP, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The terms "complementary" or "complementarity", as used herein, refer to the natural binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands and in the design and use of PNA molecules.

A "composition comprising a given polynucleotide sequence" as used herein refers broadly to any composition containing the given polynucleotide sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding NHRP (SEQ ID NOs:38-74) or fragments thereof may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., SDS) and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, has been extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or has been assembled from the overlapping sequences of more than one Incyte Clone using a computer program for fragment assembly (e.g., GELVIEW™ Fragment Assembly system, GCG, Madison, WI). Some sequences have been both extended and assembled to produce the consensus sequence .

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of a ribonucleic acid that is similar to a polynucleotide encoding an NHRP by northern analysis is indicative of the presence of mRNA encoding NHRP in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

A "deletion", as used herein, refers to a change in the amino acid or nucleotide sequence and results in the absence of one or more amino acid residues or nucleotides.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding or complementary to NHRP or the encoded NHRP. Such modifications include, for example, replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative encodes a polypeptide which retains the biological or immunological function of the natural molecule. A derivative polypeptide is one which is modified by glycosylation, pegylation, or any similar process which retains the biological or immunological function of the polypeptide from which it was derived.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or hybridization probe will compete for and inhibit the binding of a completely homologous sequence to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity). In the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

Human artificial chromosomes (HACs) are linear microchromosomes which may

contain DNA sequences of 10K to 10M in size and contain all of the elements required for stable mitotic chromosome segregation and maintenance (Harrington, J.J. et al. (1997) Nat Genet. 15:345-355).

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively, as compared to the naturally occurring molecule.

"Microarray" refers to an array (or arrangement) of distinct oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, gel, polymer, chip, glass slide, or any other suitable support.

The term "modulate", as used herein, refers to a change in the activity of NHRP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional or immunological properties of NHRP.

"NHRP" is a collective designation and represents the group of human polypeptides, NHRP-1, NHRP-2, NHRP-3, NHRP-4, NHRP-5, NHRP-6, NHRP-7, NHRP-8, NHRP-9, NHRP-10, NHRP-11, NHRP-12, NHRP-13, NHRP-14, NHRP-15, NHRP-16, NHRP-17, NHRP-18, NHRP-19, NHRP-20, NHRP-21, NHRP-22, NHRP-23, NHRP-24, NHRP-25,

NHRP-26, NHRP-27, NHRP-28, NHRP-29, NHRP-30, NHRP-31, NHRP-32, NHRP-33, NHRP-34, NHRP-35, NHRP-36, and NHRP-37, which regulate temporal, developmental, and physiological patterns of gene expression. The gene expression of an NHRP may be activated, enhanced, or repressed in the presence of biological or environmental perturbations of any kind.

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. "Fragments" are those nucleic acid sequences which are greater than 60 nucleotides than in length, and most preferably includes fragments that are at least 100 nucleotides or at least 1000 nucleotides, and at least 10,000 nucleotides in length.

The term "oligonucleotide" refers to a nucleic acid sequence of at least about 6 nucleotides to about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20 to 25 nucleotides, which can be used in PCR amplification or hybridization assays. As used herein, oligonucleotide is substantially equivalent to the terms "amplimers","primers", "oligomers", and "probes", as commonly defined in the art.

"Peptide nucleic acid", PNA as used herein, refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least five nucleotides in length linked to a peptide backbone of amino acid residues which ends in lysine. The terminal lysine confers solubility to the composition. PNAs may be pegylated to extend their lifespan in the cell where they preferentially bind complementary single stranded DNA and RNA and stop transcript elongation (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from five amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of an NHRP encompasses the full-length NHRP and fragments thereof.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding NHRP, or fragments thereof, or NHRP itself may comprise a bodily fluid, extract from a cell, chromosome, organelle, or membrane

isolated from a cell, a cell, genomic DNA, RNA, or cDNA (in solution or bound to a solid support, a tissue, a tissue print, and the like.

The terms "specific binding" or "specifically binding", as used herein, refers to that interaction between a protein or peptide and an agonist, an antibody and an antagonist. The interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) of the protein recognized by the binding molecule. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The terms "stringent conditions"or "stringency", as used herein, refer to the conditions for hybridization as defined by the nucleic acid, salt, and temperature. These conditions are well known in the art and may be altered in order to identify or detect identical or related polynucleotide sequences. Numerous equivalent conditions comprising either low or high stringency depend on factors such as the length and nature of the sequence (DNA, RNA, base composition), nature of the target (DNA, RNA, base composition), milieu (in solution or immobilized on a solid substrate), concentration of salts and other components (e.g., formamide, dextran sulfate and/or polyethylene glycol), and temperature of the reactions (within a range from about 5°C below the melting temperature of the probe to about 20°C to 25°C below the melting temperature). One or more factors be may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for

the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

A "variant" of NHRP, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

THE INVENTION

The invention is based on the discovery of human regulatory proteins (NHRP) and the polynucleotides encoding NHRP, and the use of these compositions for the diagnosis, prevention, or treatment of diseases associated with cell proliferation. Table I below shows the protein and nucleotide sequence identification numbers, Incyte Clone number, cDNA library, NCBI homolog and NCBI sequence identifier for each of the human regulatory proteins disclosed herein.

NHRP-1 (SEQ ID NO:1) was identified in Incyte Clone 187 from the U937NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus

sequence, SEQ ID NO:38, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 187 (U937NOT01), 2127894 (KIDNNOT05), and 2457706 (ENDANOT01).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:1. NHRP-1 is 260 amino acids in length and has a potential ATP/GTP binding motif at G_{234}KVIWGKY. NHRP-1 has sequence homology with Saccharomyces cerevisiae chromosome 5, GI 603365, and is associated with cDNA libraries ~~which are immortalized or cancerous and show inflammatory or immune responses.~~

NHRP-2 (SEQ ID NO:2) was first identified in Incyte Clone 2335 from the U937NOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:39, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2335 (U937NOT01), 418514 (BRSTNOT01), and 2119754 (BRSTTUT02).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:2. NHRP-2 is 153 amino acids in length and has a potential signature for regulating chromosome condensation at A_{136}AGGVGSIVRV . NHRP-2 has sequence homology with S. pombe chromosome 1, GI 1322397, and is associated with cDNA ~~libraries which are immortalized or cancerous.~~

NHRP-3 (SEQ ID NO:3) was first identified in Incyte Clone 36079 from the HUVENOB01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:40, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 36079 (HUVENOB01), 238230 (SINTNOT02), 261496 (HNT2AGT01), 348000 (THYMNOT02), 909495 (STOMNOT02), 1265067 (SYNORAT05), 1281247 (COLNNOT16), and 1337595 (COLNNOT13).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:3. NHRP-3 is 185 amino acids in length and has several potential phosphorylation sites as designated by subscripts, T_{30}DK, S_{44}KK, K_{50}KET$_{53}$ITE, S_{104}IFD, S_{126}HR, S_{150}IR, and S_{181}GK. NHRP-3 has sequence homology with Caenorhabditis elegans, GI 899244, and is associated with cDNA libraries which are immortalized or ~~cancerous and show inflammatory or immune responses.~~

NHRP-4 (SEQ ID NO:4) was first identified in Incyte Clone 82709 from the HUVESTB01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:41, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 82709 (HUVESTB01), 1399027 (BRAITUT08), and 1688789 (PROSTUT10).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:4. NHRP-4 is 106 amino acids in length and has eight potential posttranslational processing sites as indicated by subscript $Y_{36}M_{37}NL_{39}LGMIFSMC$, $W_{58}VA_{60}VYCSFISFAM$, and $M_{80}M_{81}SSFM_{85}LSISAVVMSY$. NHRP-4 has sequence homology with C. elegans, GI 860698, and is associated with cDNA libraries which are immortalized or cancerous.

NHRP-5 (SEQ ID NO:5) was first identified in Incyte Clone 313727 from the LUNGNOT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:42, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 313727 (LUNGNOT02), 353743 (RATRNOT01), 1997282 (BRSTTUT03), and 2622246 (KERANOT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:5. NHRP-5 is 166 amino acids in length and has two potential ATP/GTP binding motifs at $G_{69}TGRGGAS$ and $G_{74}ASIYGKQ$ and a cyclophilin-type peptidyl-prolyl cis-trans isomerase signature, Y48NGTKFHRIIKDFMIQGG. NHRP-5 has sequence homology with C. elegans, GI 1330343, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-6 (SEQ ID NO:6) was first identified in Incyte Clone 965366 from the BRSTNOT05 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:43, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 965366 (BRSTNOT05), 809187(LUNGNOT04), 1231614 (BRAITUT01), and 1550581(PROSNOT06).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:6. NHRP-6 is 173 amino acids in length and has a potential ATP/GTP binding motif at $G_{118}LDELGLS$. NHRP-6 has sequence homology with C.

elegans, GI 1330401, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-7 (SEQ ID NO:7) was first identified in Incyte Clone 1282071 from the COLNNOT16 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:44, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1282071 (COLNNOT16), 760942 (BRAITUT02), 767980 (LUNGNOT04), 1451163 (PENITUT01), and 1645753 (PROSTUT09).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:7. NHRP-7 is 245 amino acids in length and has a potential ATP/GTP binding site motif at G_{158}GLVHPKT and a potential signature for regulating chromosome condensation at M_{62}CVGNRHGLLV. NHRP-7 has sequence homology with S. cerevisiae, GI 1314090 , and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-8 (SEQ ID NO:8) was first identified in 1406755 from the LATRTUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:45, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1406755 (LATRTUT02), 105 (U937NOT01), 793927 (OVARNOT03), 867458 (BRAITUT03), and 1575402 and 1576873 (LNODNOT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:8. NHRP-8 is 198 amino acids in length and has a potential EGF cysteine pattern signature, C_{60}GCGVLSIGTAM and an N-6 adenine -specific DNA methylase signature , V_{123}IMNPPF. NHRP-8 has sequence homology with C. elegans, GI 559422 , and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-9 (SEQ ID NO:9) was first identified in Incyte Clone 1522948 from the BLADTUT04 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:46, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1522948 (BLADTUT04), 1289702 (BRAINOT11), 1415249 (BRAINOT12), 1508704 (LUNGNOT14), and 1821516 (GBLATUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino

acid sequence of SEQ ID NO:9. NHRP-9 is 224 amino acids in length and has a potential ATP/GTP binding motif at G_{11}GGKAGKG, and a Bzip transcription factor signature at K_{56}HGTKNKRAALQALKR. NHRP-9 has sequence homology with C. elegans, GI 868241, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-10 (SEQ ID NO:10) was first identified in Incyte Clone 1554225 from the BLADTUT04 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:47, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1554225 (BLADTUT04), 180658 (PLACNOB01), 772780 (COLNCRT01), 813242 (LUNGNOT04), 959334 (BRSTTUT03), and 1592832 (BRAINOT14).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:10. NHRP-10 is 180 amino acids in length and has a potential signature for regulating chromosome condensation at S_{104}LILENNPLVDFVEL. NHRP-10 has sequence homology with S. cerevisiae, GI 486601, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-11 (SEQ ID NO:11) was first identified in Incyte Clone 1613785 from the COLNTUT06 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:48, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1613785 (COLNTUT06), 1661626 (BRSTNOT09), and 2476089 (SMCANOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:11. NHRP-11 is 98 amino acids in length and has a potential copper binding region signature, Q_{43}VHHF, a potential immunoglobulin-Mhc signature, F_{40}TAQVHH, and a potential thiol protease motif, H_{45}HFMELCWDKCV, all essentially in the same region of the molecule. NHRP-11 has sequence homology with S. pombe, GI 1008989, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-12 (SEQ ID NO:12) was first identified in Incyte Clone 1634175 from the COLNNOT19 cDNA library using a computer search for amino acid sequence alignments. A

consensus sequence, SEQ ID NO:49, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1634175 (COLNNOT19), 685437 (UTRSNOT02), 784882 (PROSNOT05), and 816582 (OVARTUT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:12. NHRP-12 is 168 amino acids in length and has cytochrome c-like motif at K_{84}XLCHK. NHRP-12 has sequence homology with C. elegans, GI 687880, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-13 (SEQ ID NO:13) was first identified in Incyte Clone 1675954 from the BLADNOT05 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:50, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1675954 (BLADNOT05), 1282694 (COLNNOT16), 1339873 (COLNTUT03), 1809618 (PROSTUT12), and 2210710 (SINTFET03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:13. NHRP-13 is 247 amino acids in length and has a potential ATP/GTP binding motif,A_{71}EKALGKA; a potential Myb DNA-binding domain, K_{183}TRMEERRL; and a zinc finger binding motif, C_{135}FLRCCGCVFSERALKEIKAEVCH. NHRP-13 has sequence homology with C. elegans, GI 1627533, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-14 (SEQ ID NO:14) was first identified in Incyte Clone 1707463 from the DUODNOT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:51, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1707463 (DUODNOT02), 76 (U937NOT01), 1831924 (THP1AZT01), 1861727 (PROSNOT19), and 2236855 (PANCTUT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:14. NHRP-14 is 259 amino acids in length and has a potential signature for regulating chromosome condensation at A_{26}SAGVKTLLPV. NHRP-14 has sequence homology with C. elegans, GI 527429, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-15 (SEQ ID NO:15) was first identified in Incyte Clone 1720173 from the

BLADNOT06 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:52, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1720173 (BLADNOT06), 1975614 (PANCTUT02), and 2607690 (LUNGTUT07).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:15. NHRP-15 is 165 amino acids in length and has two potential phosphorylation sites, S_{13}AEE and S_{104}AED. NHRP-15 has sequence homology with C. elegans, GI 687847, and is associated with cDNA libraries which are immortalized or cancerous.

NHRP-16 (SEQ ID NO:16) was first identified in Incyte Clone 1729330 from the BRSTTUT08 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:53, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1729330 (BRSTTUT08), 891333 (STOMTUT01), and 1235735 (LUNGFET03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:16. NHRP-16 is 89 amino acids in length and has a potential growth factor receptor signature at C_{53}VNLIPQITSIYEW . NHRP-16 has sequence homology with Arabidopsis thaliana, GI 1707018, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-17 (SEQ ID NO:17) was first identified in Incyte Clone 1746646 from the STOMTUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:54, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1746646 (STOMTUT02), 727 (U937NOT01), and 2417586 (HNT3AZT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:17. NHRP-17 is 82 amino acids in length and has a potential glycosylation site at N_{31}FSI and a potential phosphorylation site at T_{35}KED. NHRP-17 has sequence homology with C. elegans, GI 1122819, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-18 (SEQ ID NO:18) was first identified in Incyte Clone 1753185 from the

LIVRTUT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:55, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1753185 (LIVRTUT01), 761463 (BRAITUT02), 1726493 (PROSNOT14), 2151755 (BRAINOT09), and 2474454 (SMCANOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:18. NHRP-18 is 259 amino acids in length and has a potential cytokine receptor signature at C_{122}VGVGLLIATLMW; an H4 DNA binding signature at G_{28}AKRY; and a G-beta repeat at L_{232}ILLYGLSLALGWNF. NHRP-18 has sequence homology with C. elegans, GI 662895, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-19 (SEQ ID NO:19) was first identified in Incyte Clone 1844162 from the COLNNOT08 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:56, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1844162 (COLNNOT08), 1355185 (LUNGNOT09), and 2473749 (THP1NOT03).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:19. NHRP-19 is 131 amino acids in length and has a potential ATP/GTP binding motif at A_{100}LVLTGNT and a potential protein splicing signature at V_{102}LTGNT. NHRP-19 has sequence homology with C. elegans, GI 733555, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-20 (SEQ ID NO:20) was first identified in Incyte Clone 1844338 from the COLNNOT08 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:57, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1844338 (COLNNOT08), 965430 (BRSTNOT05), and 1383294 (BRAITUT08).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:20. NHRP-20 is 167 amino acids in length and has potential ATP/GTP binding motifs at G_{62}TRWFGKC and V_{99}PELDGKT. NHRP-20 has sequence homology with C. elegans, GI 6656, and is associated with cDNA libraries which are

~~immortalized or cancerous and show inflammatory or immune responses.~~

NHRP-21 (SEQ ID NO:21) was first identified in Incyte Clone 1853104 from the LUNGFET03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:58, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1853104 (LUNGFET03) and 1307819 (COLNFET02).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:21. NHRP-21 is ~~96 amino acids~~ in length and has a potential protein ~~splicing signature at~~ I_{20}IVVNT. NHRP-21 has sequence homology with C. elegans, ~~GI 995857 and is associated with cDNA libraries of fetal origin.~~

NHRP-22 (SEQ ID NO:22) was first identified in Incyte Clone 1858616 from the PROSNOT18 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:59, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1858616 (PROSNOT18), 4449 (HMC1NOT01), 2073868 (ISLTNOT01), 2107812 (BRAITUT03), and 2652869 (THYMNOT04).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:22. NHRP-22 is 133 ~~amino acids~~ in length and has a potential ATP/GTP binding motif A_{116}LKDRGLS. NHRP-22 has sequence homology with C. elegans, GI 1177284, and is associated with cDNA libraries which are immortalized or cancerous and ~~show inflammatory or immune responses.~~

NHRP-23 (SEQ ID NO:23) was first identified in Incyte Clone 1969807 from the UCMCL5T01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:60, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1969807 (UCMCL5T01), 826 (U937NOT01), and 2627338 (PROSTUT12).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:23. NHRP-23 is ~~109 amino acids~~ in length and has several potential phosphorylation sites, as designated by subscripts, at S_{68}SEE, T_{93}RR, and $T_{101}S_{102}$DD, . NHRP-23 has sequence homology with C. elegans, GI 1469002, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory ~~or immune responses.~~

NHRP-24 (SEQ ID NO:24) was first identified in Incyte Clone 1971003 from the UCMCL5T01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:61, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1971003 (UCMCL5T01), 269130 (HNT2NOT01), and 487989 (HNT2AGT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:24. NHRP-24 is 138 amino acids in length and has five potential posttranslational processing sites as indicated by subscript G_{16}FGVFFLFFGM and $L_{35}A_{36}IG_{38}$NVL F_{42}VAGLAFVIGL. NHRP-24 has sequence homology with C. elegans, GI 746540, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-25 (SEQ ID NO:25) was first identified in Incyte Clone 1972328 from the UCMCL5T01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:62, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 1972328 (UCMCL5T01) and 1869477 (SKINBIT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:25. NHRP-25 is 100 amino acids in length and has potential zinc finger motifs at C_{25}FQRCVPSLHHRALDAEEEACLH and C_{45}LHSCAGKLIHSNHRL MAAYVH. NHRP-24 has sequence homology with a human tRNA pseudogene, GI 292845, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-26 (SEQ ID NO:26) was first identified in Incyte Clone 2057883 from the BEPINOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:63, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2057883 (BEPINOT01), 278794 (TESTNOT03), 305221 (HEARNOT01), 759525 (BRAITUT02), 1447663 (PLACNOT02), and 2734548 (OVARTUT04).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:26. NHRP-26 is 314 amino acids in length and has a potential ATP/GTP binding site at A_{65}YLPTGKQ and a potential Myb DNA-binding domain at

W_{55}ATGEELKV. NHRP-26 has sequence homology with S. cerevisiae, GI 1302480, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-27 (SEQ ID NO:27) was first identified in Incyte Clone 2075409 from the ISLTNOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:64, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2075409 (ISLTNOT01), 32008 (THP1NOB01), 73315 (THP1PEB01), 827713 (PROSNOT06), and 1571159 (UTRSNOT05).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:27. NHRP-27 is 140 amino acids in length and has a potential ATP/GTP binding site at A_{68}IDIHNKT. NHRP-27 has sequence homology with S. cerevisiae, GI 603277, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-28 (SEQ ID NO:28) was first identified in Incyte Clone 2095728 from the BRAITUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:65, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2095728 (BRAITUT02), 501 (U937NOT01), 002500 (U937NOT01) and 2548750 (LUNGTUT06).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:28. NHRP-28 is 125 amino acids in length and has a potential ATP/GTP binding site at A_{60}LVKPEKT. NHRP-28 has sequence homology with C. elegans, GI 1523895, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-29 (SEQ ID NO:29) was first identified in Incyte Clone 2100016 from the BRAITUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:66, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2100016 (BRAITUT02), 318737 (EOSIHET02) and 2526645 (BRAITUT21).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:29. NHRP-29 is 142 amino acids in length and has a potential

FKBP-type peptidylprolyl cis-trans isomerase signature at M_1SYMLPHLHNGWQVDQA. NHRP-29 has sequence homology with S. cerevisiae, GI 1230688, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-30 (SEQ ID NO:30) was first identified in Incyte Clone 2126751 from the KIDNNOT05 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:67, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2126751 (KIDNNOT05), 183728 (CARDNOT01), 1436704 (PANCNOT08), 1443877 (THYRNOT03), 1705143 (DUODNOT02), and 2101739 (BRAITUT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:30. NHRP-30 is 310 amino acids in length and has a potential ATP/GTP binding site at A_{274}IKNVGKQ. NHRP-30 has sequence homology with S. cerevisiae, GI 836759, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-31 (SEQ ID NO:31) was first identified in Incyte Clone 2179882 from the SININOT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:68, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2179882 (SININOT01), 757804 (BRAITUT02), 1274619 (TESTTUT02), 1576234 (LNODNOT03), and 2594101 (OVARTUT02).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:31. NHRP-31 is 209 amino acids in length and has a potential receptor protein signature at M_1KAVVQRVTRASVTVGGE. NHRP-31 has sequence homology with S. cerevisiae, GI 1431368, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-32 (SEQ ID NO:32) was first identified in Incyte Clone 2275119 from the PROSNON01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:69, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2275119 (PROSNON01) 1456127 (COLNFET02), 1759694 (PITUNOT03), and 1871122 (SKINBIT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:32. NHRP-32 is 110 amino acids in length and has four cysteine rich cytochrome c-like signatures, C_{23}EKCDG, C_{30}VICDS, C_{58}VICGG, and C_{70}KECTI. NHRP-32 has sequence homology with S. cerevisiae, GI 1230697, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-33 (SEQ ID NO:33) was first identified in Incyte Clone 2278093 from the PROSNON01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:70, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2278093 (PROSNON01), 113311 (TESTNOT01), 558866 (MPHGLPT02), 1562726 (SPLNNOT04), 1813371 (PROSTUT12), and 2311061 (NGANNOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:33. NHRP-33 is 264 amino acids in length and has two potential lipocalin signatures as indicated by subscripts, N_{149}PG$_{151}$EDEVEGLKRLM. NHRP-33 has sequence homology with C. elegans, GI 1523932, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-34 (SEQ ID NO:34) was first identified in Incyte Clone 2345426 from the TESTTUT02 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:71, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2345426 (TESTTUT02), 761916 (BRAITUT02), 1343927 (PROSNOT11), 1822317 (GBLATUT01), and 2280630 (PROSNON01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:34. NHRP-34 is 153 amino acids in length and has a potential ATP/GTP binding motif at E_{143}VVTFPGNPN. NHRP-34 has sequence homology with C. elegans, GI 1067081, and is associated with cDNA libraries which are immortalized or cancerous and show inflammatory or immune responses.

NHRP-35 (SEQ ID NO:35) was first identified in Incyte Clone 2364523 from the ADRENOT07 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:72, was derived from the extended and overlapping nucleic

acid sequences: Incyte Clones 2364523 (ADRENOT07), 148438 (FIBRNGT01), 877235 (LUNGAST01), and 2025369 (KERANOT02).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:35. NHRP-35 is ~~150~~ amino acids in length and has a potential ATP/GTP binding motif at G_{127}RRCHFKT. NHRP-35 has sequence homology with C. elegans, GI 1703579, and is associated with cDNA libraries which are immortalized or ~~cancerous and show inflammatory or immune responses.~~

NHRP-36 (SEQ ID NO:36) was first identified in Incyte Clone 2470912 from the THP1NOT03 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:73, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2470912 (THP1NOT03), 764114 (LUNGNOT04), and 2586489 (BRAITUT22).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:36. NHRP-36 is 139 amino acids in length and has a potential C-type lectin domain, C_{14}ALIFLSVYFIITLSDLECDYINARSC. NHRP-36 has sequence homology with S. cerevisiae, GI 1322550, and is associated with cDNA libraries which are ~~immortalized or cancerous and show inflammatory or immune responses.~~

NHRP-37 (SEQ ID NO:37) was first identified in Incyte Clone 2507014 from the CONUTUT01 cDNA library using a computer search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:74, was derived from the extended and overlapping nucleic acid sequences: Incyte Clones 2507014 (CONUTUT01), 1394758 (THYRNOT03), 1650580 (PROSTUT09), 2152990 (BRAINOT09), 2361374 (LUNGFET03) and 2602153 (UTRSNOT10).

~~In one embodiment, the invention encompasses a polypeptide comprising the amino~~ acid sequence of SEQ ID NO:37. NHRP-37 is 350 amino acids in length and has two potential glycosylation sites at N_{147}TYQ and N_{185}CTQ, and several potential phosphorylation sites as designated by subscripts, S_9LND, S_{17}FAE, T_{80}WKE, T_{122}SR, S_{171}AR, T_{174}GNE, T_{187}QK, T_{237}MID, S_{293}HHD, S_{313}NT$_{315}$ER, S_{329}HLE, S_{340}ETD, and T_{342}DRD. NHRP-37 has sequence homology with S. cerevisiae, GI 1322869, and is associated with cDNA libraries ~~which are immortalized or cancerous and show inflammatory or immune responses.~~

The invention also encompasses NHRP variants which retain the biological or functional activity of NHRP. A preferred NHRP variant is one having at least 80%, and more preferably 90%, amino acid sequence identity to the NHRP amino acid sequence. A most preferred NHRP variant is one having at least 95% amino acid sequence identity to an NHRP disclosed herein (SEQ ID NOs:1-37).

The invention also encompasses polynucleotides which encode NHRP. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of NHRP can be used to produce recombinant molecules which express NHRP. In a particular embodiment, the invention encompasses a polynucleotide consisting of a nucleic acid sequence selected from the group consisting of SEQ ID NOs:38-74

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding NHRP, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring NHRP, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode NHRP and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring NHRP under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding NHRP or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding NHRP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or fragments thereof, which encode NHRP and its derivatives, entirely by synthetic chemistry. After production,

the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding NHRP or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NOs:38-74, under various conditions of stringency as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511).

Methods for DNA sequencing which are well known and generally available in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE Amplification System marketed by Gibco/BRL (Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI Catalyst and 373 and 377 DNA Sequencers (Perkin Elmer).

The nucleic acid sequences encoding NHRP may be extended utilizing a partial nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers

may be designed using commercially available software such as OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown fragment of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary

electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode NHRP may be used in recombinant DNA molecules to direct expression of NHRP, fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced, and these sequences may be used to clone and express NHRP.

As will be understood by those of skill in the art, it may be advantageous to produce NHRP-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce an RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter NHRP encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding NHRP may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of NHRP activity, it may be useful to encode a chimeric NHRP protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the NHRP encoding sequence and the heterologous protein sequence, so that NHRP may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding NHRP may be synthesized, in whole or

in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of NHRP, or a fragment thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of NHRP, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active NHRP, the nucleotide sequences encoding NHRP or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding NHRP and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding NHRP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with

virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. The invention is not limited by the host cell employed.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding NHRP, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for NHRP. For example, when large quantities of NHRP are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding NHRP may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by

adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding NHRP may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express NHRP. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding NHRP may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of NHRP will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which NHRP may be expressed (Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding NHRP may

be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing NHRP in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained and expressed in a plasmid. HACs of 6 to 10M are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding NHRP. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding NHRP, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38), are available from the American Type Culture Collection (ATCC;

Bethesda, MD) and may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express NHRP may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131). .

Although the presence/absence of marker gene expression suggests that the gene of

interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding NHRP is inserted within a marker gene sequence, transformed cells containing sequences encoding NHRP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding NHRP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding NHRP and express NHRP may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding NHRP can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding NHRP. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding NHRP to detect transformants containing DNA or RNA encoding NHRP.

A variety of protocols for detecting and measuring the expression of NHRP, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on NHRP is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding NHRP include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding NHRP, or any fragments

thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used for ease of detection, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding NHRP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode NHRP may be designed to contain signal sequences which direct secretion of NHRP through a prokaryotic or eukaryotic cell membrane. Other constructions may be used to join sequences encoding NHRP to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and NHRP may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing NHRP and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3: 263-281) while the enterokinase cleavage site provides a means for purifying NHRP from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

In addition to recombinant production, fragments of NHRP may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of NHRP may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exits among the human regulatory proteins of the invention. The expression of NHRP is closely associated with cell proliferation. Therefore, in cancers or immune disorders where NHRP is an activator, transcription factor, or enhancer, is being expressed, and is promoting cell proliferation; it is desirable to decrease the expression of NHRP. In cancers where NHRP is an inhibitor or suppressor and is not controlling cell proliferation, it is desirable to provide the protein or increase the expression of NHRP.

In one embodiment, where expression of NHRP is decreased, NHRP or a fragment or derivative thereof may be administered to a subject to prevent or treat cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, an agonist which is specific for NHRP may be administered to a subject to prevent or treat cancer, including but limited to those cancers listed above.

In another further embodiment, a vector capable of expressing NHRP, or a fragment or a derivative thereof, may be administered to a subject to prevent or treat cancer including, but not limited to, those cancers listed above.

In a further embodiment, antagonists which decrease the expression and activity of NHRP may be administered to a subject to prevent or treat cancer such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, and teratocarcinoma. Such cancers

include, but are not limited to, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. In one aspect, antibodies which specifically bind NHRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express NHRP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding NHRP may be administered to a subject to treat or prevent cancer including, but not limited to, those cancers listed above.

In one embodiment, antagonists which decrease the activity of NHRP may be administered to a subject to prevent or treat immune responses. Such responses may be associated with AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and autoimmune thyroiditis; complications of cancer, hemodialysis, extracorporeal circulation; viral, bacterial, fungal, parasitic, protozoal, and helminthic infections and trauma. In one aspect, antibodies which specifically bind NHRP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express NHRP.

In another embodiment, a vector expressing the complement of the polynucleotide encoding NHRP may be administered to a subject to treat or prevent immune responses including, but not limited to, those listed above

In one further embodiment, NHRP or a fragment or derivative thereof may be added to cells to stimulate cell proliferation. In particular, NHRP may be added to a cell in culture or cells *in vivo* using delivery mechanisms such as liposomes, viral based vectors, or electroinjection for the purpose of promoting cell proliferation and tissue or organ regeneration. Specifically, NHRP may be added to a cell, cell line, tissue or organ culture *in*

vitro or *ex vivo* to stimulate cell proliferation for use in heterologous or autologous transplantation. In some cases, the cell will have been preselected for its ability to fight an infection or a cancer or to correct a genetic defect in a disease such as sickle cell anemia, β thalassemia, cystic fibrosis, or Huntington's chorea.

In another embodiment, an agonist which is specific for NHRP may be administered to a cell to stimulate cell proliferation, as described above.

In another embodiment, a vector capable of expressing NHRP, or a fragment or a derivative thereof, may be administered to a cell to stimulate cell proliferation, as described above.

In other embodiments, any of the therapeutic proteins, antagonists, antibodies, agonists, complementary sequences or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of NHRP may be produced using methods which are generally known in the art. In particular, purified NHRP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind NHRP.

Antibodies to NHRP may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with NHRP or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active

substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are espécially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to NHRP have an amino acid sequence consisting of at least five amino acids and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of NHRP amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to NHRP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce NHRP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for NHRP may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between NHRP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering NHRP epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding NHRP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, the complement of the polynucleotide encoding NHRP may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding NHRP. Thus, complementary molecules or fragments may be used to modulate NHRP activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligonucleotides or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding NHRP.

Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct vectors which will express nucleic acid sequence which is complementary to the polynucleotides of the gene encoding NHRP. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra).

Genes encoding NHRP can be turned off by transforming a cell or tissue with

expression vectors which express high levels of a polynucleotide or fragment thereof which encodes NHRP. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing complementary sequences or antisense molecules (DNA, RNA, or PNA) to the control, 5' or regulatory regions of the gene encoding NHRP (signal sequence, promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, NY). The complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding NHRP.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by

testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding NHRP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections or polycationic amino polymers (Goldman, C.K. et al. (1997) Nature Biotechnology 15:462-66; incorporated herein by reference) may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may

consist of NHRP, antibodies to NHRP, mimetics, agonists, antagonists, or inhibitors of NHRP. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium

alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers may also be used for delivery. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with

many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of NHRP, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example NHRP or fragments thereof, antibodies of NHRP, agonists, antagonists or inhibitors of NHRP, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind NHRP may be used for the diagnosis of conditions or diseases characterized by expression of NHRP, or in assays to monitor patients being treated with NHRP, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for NHRP include methods which utilize the antibody and a label to detect NHRP in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring NHRP are known in the art and provide a basis for diagnosing altered or abnormal levels of NHRP expression. Normal or standard values for NHRP expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to NHRP under conditions suitable for complex formation The amount of standard complex formation may be quantified by various methods, but preferably by photometric,

means. Quantities of NHRP expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding NHRP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of NHRP may be correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of NHRP, and to monitor regulation of NHRP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding NHRP or closely related molecules, may be used to identify nucleic acid sequences which encode NHRP. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding NHRP, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the NHRP encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NOs:38-74 or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring NHRP.

Means for producing specific hybridization probes for DNAs encoding NHRP include the cloning of nucleic acid sequences encoding NHRP or NHRP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin

coupling systems, and the like.

Polynucleotide sequences encoding NHRP may be used for the diagnosis of conditions, disorders, or diseases which are associated with expression of NHRP. Examples of such conditions or diseases include adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and cancers of the adrenal gland, bladder, bone, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, bone marrow, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus; and immune disorders such as AIDS, Addison's disease, adult respiratory distress syndrome, allergies, anemia, asthma, atherosclerosis, bronchitis, cholecystitus, Crohn's disease, ulcerative colitis, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, atrophic gastritis, glomerulonephritis, gout, Graves' disease, hypereosinophilia, irritable bowel syndrome, lupus erythematosus, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, rheumatoid arthritis, scleroderma, Sjögren's syndrome, and thyroiditis. The polynucleotide sequences encoding NHRP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dipstick, pin, ELISA assays or microarrays utilizing fluids or tissues from patient biopsies to detect altered NHRP expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding NHRP may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding NHRP may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding NHRP in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring

the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of NHRP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes NHRP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding NHRP may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'->3') and another with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of NHRP include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides derived from any of the polynucleotide sequences described herein may be used in microarrays. The microarrays are used to monitor the expression level of large numbers of genes simultaneously (to produce a transcript image), and to identify genetic variants, mutations and polymorphisms. This information will be useful in determining gene function, understanding the genetic basis of disease, diagnosing disease, and in developing and in monitoring the activities of therapeutic agents.

In one embodiment, the microarray is prepared and used according to the methods described in PCT application WO95/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference.

The microarray is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. Microarrays may contain oligonucleotides which cover the known 5', or 3', sequence, or contain sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray may be oligonucleotides that are specific to a gene or genes of interest in which at least a fragment of the sequence is known or that are specific to one or more unidentified cDNAs which are common to a particular cell type, developmental or disease state.

In order to produce oligonucleotides to a known sequence for a microarray, the gene of interest is examined using a computer algorithm which starts at the 5' or more preferably at the 3' end of the nucleotide sequence. The algorithm identifies oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. The oligomers are

synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, the oligomers may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application WO95/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array may be produced by hand or using available devises (slot blot or dot blot apparatus) materials and machines (including robotic instruments) and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots, or any other multiple which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using the microarray, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray so that the probe sequences hybridize to complementary oligonucleotides of the microarray. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, mutations, variants, or polymorphisms among samples.

In another embodiment of the invention, the nucleic acid sequences which encode NHRP may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular

chromosome, to a specific region of a chromosome or to artificial chromosome constructions, such as human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C.M. (1993) Blood Rev. 7:127-134, and Trask, B.J. (1991) Trends Genet. 7:149-154.

Fluorescent in situ hybridization (FISH as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in various scientific journals or at Online Mendelian Inheritance in Man (OMIM). Correlation between the location of the gene encoding NHRP on a physical chromosomal map and a specific disease , or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, NHRP, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be

free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between NHRP and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to NHRP large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with NHRP, or fragments thereof, and washed. Bound NHRP is then detected by methods well known in the art. Purified NHRP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding NHRP specifically compete with a test compound for binding NHRP. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with NHRP.

In additional embodiments, the nucleotide sequences which encode NHRP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

EXAMPLES

For purposes of example, the preparation and sequencing of the UCMCL5T01 cDNA library, from which Incyte Clones 1969807, 1971003, and 1972328 were isolated, is described. Preparation and sequencing of cDNAs in libraries in the LIFESEQ™ database have varied slightly over time and the gradual changes involved use of particular kits, plasmids, and machinery available at the particular time the library was made and analyzed.

I UCMCL5T01 cDNA Library Construction

The UCMCL5T01 cDNA library was constructed from mRNA isolated from the lysates of umbilical cord mononuclear cells pooled from 12 individuals following culture with IL-5 for five days. The cells were homogenized and lysed using a Brinkmann Homogenizer Polytron PT-3000 (Brinkmann Instruments, Westbury, NJ) in guanidinium isothiocyanate solution. The lysate was centrifuged over a 5.7 M CsCl cushion using an Beckman SW28 rotor in a Beckman L8-70M Ultracentrifuge (Beckman Instruments) for 18 hours at 25,000 rpm at ambient temperature. The RNA was extracted with acid phenol pH 4.7, precipitated using 0.3 M sodium acetate and 2.5 volumes of ethanol, resuspended in RNAse-free water, and DNase treated at 37°C. The mRNA was then isolated with the Qiagen Oligotex kit (QIAGEN, Inc., Chatsworth, CA) and used to construct the cDNA library.

The mRNA was handled according to the recommended protocols in the SuperScript Plasmid System for cDNA synthesis and plasmid cloning (Cat. #18248-013, Gibco/BRL). cDNAs were fractionated on a Sepharose CL4B column (Cat. #275105-01, Pharmacia), and those cDNAs exceeding 400 bp were ligated into pINCYI. The plasmid pINCYI was subsequently transformed into DH5a™ competent cells (Cat. #18258-012, Gibco/BRL).

II Isolation and Sequencing of cDNA Clones

Plasmid DNA was released from the cells and purified using the REAL Prep 96 plasmid kit (Catalog #26173, QIAGEN, Inc.). This kit enabled the simultaneous purification of 96 samples in a 96-well block using multi-channel reagent dispensers. The recommended protocol was employed except for the following changes: 1) the bacteria were cultured in 1 ml of sterile Terrific Broth (Catalog #22711, Gibco/BRL) with carbenicillin at 25 mg/L and glycerol at 0.4%; 2) after inoculation, the cultures were incubated for 19 hours and at the end of incubation, the cells were lysed with 0.3 ml of lysis buffer; and 3) following isopropanol precipitation, the plasmid DNA pellet was resuspended in 0.1 ml of distilled water. After the last step in the protocol, samples were transferred to a 96-well block for storage at 4° C.

The cDNAs were sequenced by the method of Sanger et al. (1975, J. Mol. Biol. 94:441f), using a Hamilton Micro Lab 2200 (Hamilton, Reno, NV) in combination with

Peltier Thermal Cyclers (PTC200 from MJ Research, Watertown, MA) and Applied Biosystems 377 DNA Sequencing Systems; and the reading frame was determined.

Most of the sequences disclosed herein were sequenced according to standard ABI protocols, using ABI kits (Cat. Nos. 79345, 79339, 79340, 79357, 79355). The solution volumes were used at 0.25x - 1.0x concentrations. Some of the sequences disclosed herein were sequenced using different solutions and dyes which, unless otherwise noted, came from Amersham Life Science (Cleveland, OH). The samples were loaded on wells in volumes of 2 μL per well for sequencing in ABI sequencers.

III Homology Searching of cDNA Clones and Their Deduced Proteins

The nucleotide sequences and/or amino acid sequences of the Sequence Listing were used to query sequences in the GenBank, SwissProt, BLOCKS, and Pima II databases. These databases, which contain previously identified and annotated sequences, were searched for regions of homology using BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol 36:290-300; Altschul, et al. (1990) J. Mol. Biol. 215:403-410).

BLAST produced alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST was especially useful in determining exact matches or in identifying homologs which may be of prokaryotic (bacterial) or eukaryotic (animal, fungal, or plant) origin. Other algorithms such as the one described in Smith, T. et al. (1992, Protein Engineering 5:35-51), incorporated herein by reference, could have been used when dealing with primary sequence patterns and secondary structure gap penalties. The sequences disclosed in this application have lengths of at least 49 nucleotides, and no more than 12% uncalled bases (where N is recorded rather than A, C, G, or T).

The BLAST approach searched for matches between a query sequence and a database sequence. BLAST evaluated the statistical significance of any matches found, and reported only those matches that satisfy the user-selected threshold of significance. In this application, threshold was set at 10^{-25} for nucleotides and 10^{-14} for peptides.

Incyte nucleotide sequences were searched against the GenBank databases for primate

(pri), rodent (rod), and other mammalian sequences (mam); and deduced amino acid sequences from the same clones were then searched against GenBank functional protein databases, mammalian (mamp), vertebrate (vrtp), and eukaryote (eukp) for homology. The relevant database for a particular match were reported as GIxxx±p (where xxx is pri, rod, etc., and if present, p = peptide).

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., supra).

Analogous computer techniques use BLAST to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding NHRP occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V Extension of NHRP Encoding Polynucleotides

The nucleic acid sequence of an Incyte Clone disclosed in the Sequence Listing was used to design oligonucleotide primers for extending a partial nucleotide sequence to full length. One primer was synthesized to initiate extension in the antisense direction, and the other was synthesized to extend sequence in the sense direction. Primers were used to facilitate the extension of the known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers were designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be about 22 to about 30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures of about 68°to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries (Gibco/BRL) were used to extend the sequence If more than one extension is necessary or desired, additional sets of primers are designed to further extend the known region.

High fidelity amplification was obtained by following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR was performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10 μl aliquot of the reaction mixture was analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were

successful in extending the sequence. Bands thought to contain the largest products were excised from the gel, purified using QIAQuick™ (QIAGEN Inc., Chatsworth, CA), and trimmed of overhangs using Klenow enzyme to facilitate religation and cloning.

After ethanol precipitation, the products were redissolved in 13 μl of ligation buffer, 1μl T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase were added, and the mixture was incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) were transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook et al., supra). After incubation for one hour at 37° C, the E. coli mixture was plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies were randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture was transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample was transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction were added to each well. Amplification was performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions were run on agarose gels together with molecular weight markers. The sizes of the PCR products were compared to the original partial cDNAs, and appropriate clones were selected, ligated into plasmid, and sequenced.

In like manner, the nucleotide sequence of SEQ ID NO:38-74 are used to obtain 5' regulatory sequences using the procedure above, oligonucleotides designed for 5' extension, and an appropriate genomic library.

VI Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NOs:38-74 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 superfine resin column (Pharmacia & Upjohn). A aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours, hybridization patterns are compared visually.

VII Microarrays

To produce oligonucleotides for a microarray, SEQ ID Nos:38-74 were examined using a computer algorithm which starts at the 3' end of the nucleotide sequence. The algorithm identified oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that would interfere with hybridization. The algorithm identified approximately 20 sequence-specific oligonucleotides of 20 nucleotides in length (20-mers). A matched set of oligonucleotides was created in which one nucleotide in the center of each sequence was altered. This process was repeated for each gene in the microarray, and double sets of twenty 20 mers were synthesized and arranged on the surface of the silicon chip using a light-

directed chemical process (Chee, M. et al., PCT/WO95/11995, incorporated herein by reference).

In the alternative, a chemical coupling procedure and an ink jet device were used to synthesize oligomers on the surface of a substrate (Baldeschweiler, J.D. et al., PCT/WO95/25116, incorporated herein by reference). In another alternative, a "gridded" array analogous to a dot (or slot) blot was used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. A typical array may be produced by hand or using available materials and machines and contain grids of 8 dots, 24 dots, 96 dots, 384 dots, 1536 dots or 6144 dots. After hybridization, the microarray was washed to remove nonhybridized probes, and a scanner was used to determine the levels and patterns of fluorescence. The scanned images were examined to determine degree of complementarity and the relative abundance/expression level of each oligonucleotide sequence in the micro-array.

VIII Complementary Polynucleotides

Sequence complementary to the sequence encoding NHRP, or any part thereof, is used to detect, decrease or inhibit expression of naturally occurring NHRP. Although use of oligonucleotides comprising from about 15 to about 30 base-pairs is described, essentially the same procedure is used with smaller or larger sequence fragments. Appropriate oligonucleotides are designed using Oligo 4.06 software and the coding sequence of NHRP, SEQ ID NOs:38-74. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the transcript encoding NHRP.

IX Expression of NHRP

Expression of NHRP is accomplished by subcloning the cDNAs into appropriate vectors and transforming the vectors into host cells. In this case, the cloning vector is also used to express NHRP in E. coli. Upstream of the cloning site, this vector contains a

promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of NHRP into the bacterial growth media which can be used directly in the following assay for activity.

X Demonstration of NHRP Activity

NHRP can be expressed in a mammalian cell line such as DLD-1 or HCT116 (ATCC; Bethesda, MD) by transforming the cells with a eukaryotic expression vector encoding NHRP. Eukaryotic expression vectors are commercially available and the techniques to introduce them into cells are well known to those skilled in the art. The effect of NHRP on cell morphology may be visualized by microscopy; the effect on cell growth may be determined by measuring cell doubling-time; and the effect on tumorigenicity may be assessed by the ability of transformed cells to grow in a soft agar growth assay (Groden, J. et al. (1995) Cancer Res. 55:1531-1539).

XI Production of NHRP Specific Antibodies

NHRP that is substantially purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from SEQ ID NOs:38-74 is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (supra), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole

limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al., supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio iodinated, goat anti-rabbit IgG.

XII Purification of Naturally Occurring NHRP Using Specific Antibodies

Naturally occurring or recombinant NHRP is substantially purified by immunoaffinity chromatography using antibodies specific for NHRP. An immunoaffinity column is constructed by covalently coupling NHRP antibody to an activated chromatographic resin, such as CNBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing NHRP is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of NHRP (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/protein binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and NHRP is collected.

XIII Identification of Molecules Which Interact with NHRP

NHRP or biologically active fragments thereof are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133: 529). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled NHRP, washed and any wells with labeled NHRP complex are assayed. Data obtained using different concentrations of NHRP are used to calculate values for the number, affinity, and association of NHRP with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the

scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.